











A CROSS NATIONAL COMPARISON OF THE RELIABILITY OF  
SELECTED MEASUREMENTS FROM CONSUMER SURVEYS\*

Harry L. Davis\*\*, Susan P. Douglas\*\*\*,  
and Alvin J. Silk\*\*\*\*

WP# 1110-80

April 1980

**THE MARKETING CENTER**

Massachusetts Institute of Technology  
Alfred P. Sloan School of Management  
50 Memorial Drive  
Cambridge, Massachusetts 02139  
(617) 253-6615



A CROSS NATIONAL COMPARISON OF THE RELIABILITY OF  
SELECTED MEASUREMENTS FROM CONSUMER SURVEYS\*

Harry L. Davis\*\*, Susan P. Douglas\*\*\*,  
and Alvin J. Silk\*\*\*\*

WP# 1110-80

April 1980

\* The authors are indebted to Michael Baker, Peter Doyle, Jean-Marie Lefebvre, and Benny Rigaux-Bricmont for their efforts in collecting the data upon which this study is based. Thanks are due to Alain Bultez for data processing assistance and to Richard Bagozzi and William Wells for helpful comments. The financial support of the European Institute for Advanced Studies in Management and the Marketing Science Institute is gratefully acknowledged. A version of this paper was presented at the American Marketing Association's Attitude Research Conference, Carlsbad, California, March 2-5, 1980.

\*\* Graduate School of Business, University of Chicago.

\*\*\* Graduate School of Business Administration, New York University.

\*\*\*\* Sloan School of Management, Massachusetts Institute of Technology.



## ABSTRACT

Cross national surveys are frequently undertaken for purposes of making comparisons among markets with respect to factors relevant to marketing policies. A potentially troublesome threat to the validity of conclusions drawn from such comparisons is that due to differences in the reliability of measurements arising from linguistic and conceptual non-equivalencies of questionnaire instruments. This paper investigates the question of whether the reliability of measures commonly used in marketing surveys differs cross nationally. Results are presented from a five country study. Significant between-sample reliability differentials were uncovered but no systematic tendency for particular national samples or language groups to exhibit consistently high or low reliability across different types of measures and variables was observed. Between-sample reliability differentials appeared less likely to occur for measures of "hard" variables (e.g. demographics) than for measures of "soft" variables (e.g., life style/attitudinal variables). Implications of these findings for the design and analysis of cross national surveys are discussed.



## INTRODUCTION

Multinational firms are often confronted with the need to carry out comparable marketing surveys with respondents located in several different countries. The predominant concern in such investigations is, of course, with cross-national comparisons. Whether a cross-national study concludes that markets are "similar" or "different" has far-reaching policy implications. The discovery of country-by-country differences provides a basis for developing separate "localized" marketing strategies, while the absence of such differences bolsters the case for "standardized" marketing programs (Buzzell, 1968).

A key issue that invariably arises in connection with these multinational research efforts is whether observed similarities or differences between markets are, in fact, real. Few experienced in primary data collection would question the proposition that the possible threats to validity (Cook and Campbell, 1979) increase dramatically as the number and diversity of countries encompassed by a consumer research project is expanded. A large body of research has accumulated on the problems posed by various sources of systematic and random errors which afflict marketing studies (Brown, 1969, and Farley and Howard, 1975). For several years now Lipstein (1975a, 1975b, 1977) has been emphasizing to the marketing research community here and abroad that the field's preoccupation with sampling error is misdirected when one examines the magnitude of non-sampling error present in survey results, which, symptomatically, is only very rarely estimated. While sample sizes are increased to diminish sampling error, Lipstein argues that practices followed in accomplishing this expansion are such that the expected gains due to reductions in sampling error may be offset by increases in non-sampling

error and a kind of diseconomy of scale holds for the overall precision of survey measurements. Although Lipstein has not explicitly discussed cross-national research in this context, his agreement would appear to hold in spades for the international area.

The problem of non-sampling error has certainly not gone unrecognized in the literature on cross-cultural and cross-national research, but the emphasis tends to be on procedures for controlling and minimizing measurement error rather than on diagnosing and estimating its magnitude empirically, much less coping with its inevitable presence. Much has been written about the need to insure linguistic equivalence of items used in cross-cultural studies, and various back translation schemes have been proposed to facilitate achieving same (Whiting, 1968). To illustrate, translating "bakery" into French, one needs to be mindful of the distinction between a "boulangerie" and a "patisserie".

Researchers must also pay attention to conceptual equivalence. The same words, properly translated, may have varying connotations in different countries. Rao and Rao (1979), for example, have demonstrated that the dimensionality of a well-known scale of familism differs between Canadian and Indian respondents due to a broader conception of the family unit in Indian culture. In the literature on cross-cultural research, discussions of conceptual equivalence contrast "emic" and "etic" studies:

The emic approach attempts to obtain the best possible description of a phenomenon occurring in a particular local population by utilizing concepts employed in that population. It is allegedly the most accurate description of a phenomenon. However, emic data cannot be compared across cultures because, by definition, the concepts developed in a single culture may not be universal. The etic approach studies a phenomenon by utilizing concepts with generality beyond a single local population (Malpass, 1977, pp. 1074-75).

Most cross-national marketing surveys, including the one reported here, can probably be best characterized as modified etic studies since general questionnaire items are employed with some limited adaptations made to accommodate local populations.

Linguistic and conceptual non-equivalencies are sources of non-sampling error in cross-national studies and their presence will produce differentials in the biases or in the reliabilities of the measurements across the non-equivalent groups. Wherever possible, steps should be taken to reduce non-equivalencies and their adverse effects on reliability. In a marketing research context, Angelmar and Pras (1978) have shown how adjectives used to define response scale alternatives can be developed in different languages to attain metrically equivalent scale values and homogeneity of meaning. Clearly, such efforts are valuable, and more are needed. However, the development of improved cross-cultural methods is a slow process, and in the end some degree of fallibility will always remain in our measurements. As Mayer (1978, p. 77) puts it:

In order to have truly comparable data from multinational markets, it is not sufficient to attempt to control sources of error. Since such error will never be completely eliminated, the data, having been collected, have to be adjusted for residual error sources. However, before such adjustments can be done, much more knowledge is required.

One type of knowledge that is clearly needed is that which comes from identifying the presence of non-sampling error and estimating its relative magnitude. The unreliability of instruments looms as an especially troublesome but largely neglected source of non-sampling error and threat to the validity of cross-national studies. As is stressed in the literature of econometrics, psychometrics, and statistics, unreliability of observations as a consequence of the presence of random measurement error

attenuates the precision of estimators, and reduces the power of statistical tests of hypotheses (Cochran, 1968). Given that questions of how markets compare are paramount in cross-national studies, such effects of errors-in-the-variables take on an added significance because of the possibility of varying levels of reliability across countries. What might first appear to be a cross-national difference in, say, the relationship between some predictor and criterion variable observed in a segmentation study, instead might turn out to be solely a reflection of variations in the reliability of the underlying measurements employed in the analysis.

This paper explores the question of whether the reliability of measures obtained from marketing surveys differs cross-nationally. Results are reported from similar studies conducted in three French-speaking countries and two English-language countries. Cross-national reliability comparisons are made for three types of measurements commonly employed in consumer research: demographic and other background characteristics, reports of involvement in household tasks and decisions, and life style variables. Information pertaining to background characteristics and involvement in household tasks and decisions is typically used either to screen respondents and determine their eligibility for participation in a study or to define classificatory/explanatory variables and co-variates employed in some form of subsequent statistical analysis. Life style/psychographic variables have received much attention in recent years as classificatory or explanatory variables in segmentation studies. For the background characteristics and reports of involvement in household tasks and decisions, we assess the degree of convergence in measures of the variables obtained from husbands and wives in the same families. In the case of the involvement ratings, we also examine their discriminant ability. Finally, the internal

consistency of composite scores for a set of life style variables is investigated.

In the next section we describe the study's data collection procedures, the variables investigated, and the reliability indices employed. Following this, we present the cross-sample reliability comparisons for the three types of measurements. Finally, the results are summarized and their implications discussed.

## METHOD

### Data Collection

The data to be discussed were collected as part of a larger study of family decision making conducted in five countries: U.S., Great Britain, France, Belgium, and Canada. Convenience samples of households were selected from among residents of urban communities in each country. An unusual feature of this investigation compared to most marketing research studies and one that allows certain types of reliability assessments discussed below to be made, was that measurements were obtained by means of self-administered questions, from both husbands and wives in the same families. The names of the cities where respondents resided and the sample sizes per city in terms of households (2 respondents/household) were as follows: Chicago (161), London/Glasgow (95), Paris (109), Brussels (76), and Quebec (91). The samples are somewhat upscale in education and income as compared to the populations of the communities where they resided except in the case of Quebec. Demographic profiles of the five samples are presented in Douglas (1979, p. 370).

The questionnaire was originally written in English and then translated into French. While Brussels, Paris, and Quebec are all French-speaking communities, there are important differences among those cities in the language used. To deal with this, the initial French version of the questionnaire was modified by "local experts" to reflect the language usage habits in each area. Each version of the questionnaire so adapted was then translated back into English and checked for consistency with the original.

Here we examine data for three types of measures, selected to span the diversity of constructs and levels of measurement commonly employed in consumer surveys: (1) demographic and background characteristics; (2) self-reports of behavior in the form of ratings about involvement in household tasks and decisions; and (3) life-style/psychographic variables. Table 1 lists the specific variables considered for each type of measure. The first two types of variables were measured on single item categorical scales while the life-style variables are composite scores, obtained by summing responses to multiple items which purport to measure the same underlying construct.

-----

INSERT TABLE 1 HERE

-----

The ten background characteristics considered below included demographic and socio-economic variables as well as some straightforward household descriptors (type of dwelling unit and appliance ownership). For each of the five household tasks/decisions covered by the questionnaire, respondents were asked to rate their involvement relative to that of their spouses on a three-point scale: "mainly husband", "joint or shared responsibility", and "mainly wife".

The six life style variables consisted of two types. The first were concerned with three dimensions of marital and sex role attitudes, labelled: "male dominance", "task allocation", and "female role perceptions". The male dominance measure consisted of four items taken from a scale developed by Hoffman (1960) while the task allocation (5 items) and female role perception (4 items) measures were developed in previous work by Davis (1972). The second group of variables were three measures of individual traits or value/personality orientations termed "orderliness", "anxiety and control" and "traditionalism" which are discussed in Douglas and Wind (1978). All six of these measures have been used in earlier studies but primarily with U.S. samples. A copy of the complete questionnaire is available upon request from the first author. Questionnaires covering all these measures were administered separately to husbands and wives from the same families to avoid giving them the opportunity to interact with one another in the course of filling them out.

#### Reliability Assessment

The reliability of the family background characteristics and involvement ratings was assessed by determining the degree of consistency between husbands' and wives' reports of these variables, each of which was measured by a single questionnaire item consisting of a categorical scale. To accomplish this, we cross tabulated the two sets of responses and computed a measure of agreement between them. The coefficient used here is that proposed by Cohen (1960) and discussed in Bishop, Feinberg, and Holland (1975, pp. 395-397). As applied here, the coefficient is defined as follows:

$$\kappa = \frac{O - E}{1 - E},$$

where:

O = observed proportion of cases within the sample where the husband and wife from the same family gave identical responses in reporting a particular variable.

E = expected proportion of identical responses under the assumption that the husband's and wife's reports are statistically independent of one another.

The coefficient,  $\kappa$ , reflects the excess of observed over chance agreement, normalized by the maximum possible value of this difference, given the particular form of the marginal distribution observed for the two sets of responses. The coefficient is zero when the observed agreement is just equal to that expected by chance and unity when the maximum possible excess of observed over chance agreement is obtained. Negative values of  $\kappa$  indicate less observed agreement than expected by chance. Note that this coefficient is a measure of agreement which depends only on the frequency of identical responses (represented by entries in the main diagonal of the contingency table formed by cross-tabulating husbands' and wives' responses to the same question) as distinct from an index of association or correlation which would also take account of non-identical responses (off-diagonal frequencies). Fleiss (1975) examined the numerous measures of agreement for categorical data which have been proposed in the psychometric and statistics literature and concluded that  $\kappa$  is one of only two measures "defensible both as chance-corrected measures and as intraclass correlation coefficients." More recently Kraemer (1979) has shown how  $\kappa$  relates to the classical psychometric model for reliability of interval data (the ratio of true score variance to observed score variance) and discussed its interpretation and use "to indicate the degree of loss of precision or power of statistical procedures" due to the unreliability of observations.

Previous applications of  $\kappa$  to marketing research problems similar to the present one may be found in Davis and Ragsdale (1979) and Silk and Kalwani (1979).

The calculation of  $\kappa$  may be illustrated with the data for the following contingency table, formed by cross-tabulating husbands' and wives' reports of family income for the Chicago sample. Respondents indicated their income level by selecting one of six categories defined by a particular range of monetary values.

|                    |     | Wives' Response |            |             |              |             |              | Total |
|--------------------|-----|-----------------|------------|-------------|--------------|-------------|--------------|-------|
|                    |     | I               | II         | III         | IV           | V           | VI           |       |
| Husbands' Response | I   | (0)             |            |             |              |             |              | 0     |
|                    | II  |                 | 3<br>(0.1) | 2           |              |             |              | 5     |
|                    | III |                 |            | 19<br>(2.8) | 1            |             |              | 20    |
|                    | IV  |                 |            | 1           | 35<br>(1.07) | 3           |              | 39    |
|                    | V   |                 |            |             | 6            | 26<br>(6.5) | 2            | 34    |
|                    | VI  |                 |            |             | 1            | 1           | 57<br>(22.2) | 59    |
| Total              |     | 0               | 3          | 22          | 42           | 30          | 49           | 157   |

The cell entries are the observed frequencies and the figures in parenthesis along the main diagonal are expected frequencies (for the main diagonal cells) computed using the observed marginal distributions of the two sets of responses shown in the row and column totals. Applying these data to the computational formula for  $\kappa$  cited above we obtain:

$$O = \frac{1}{157} (0 + 3 + 19 + 35 + 26 + 57) = .892$$

$$E = \frac{1}{157} (0 + 0.1 + 2.8 + 10.7 + 6.5 + 22.2) = .269$$

$$K = \frac{.892 - .269}{1 - .269} = +.852$$

A different procedure is needed to assess the reliability of individual-difference or life-style variables where the measures are composite scores, i.e., a total score was obtained for each individual by summing his/her responses to the set of separate items comprising the scale. The Kuder-Richardson (Formula 20) coefficient is an index of reliability for composite scores where the response scale to the individual items is a dichotomous one (Guilford, 1954, pp. 380-383). It is a special case of Cronbach's  $\alpha$  reliability coefficient for interval data (Cronbach, 1951) and represents a lower bound on the population reliability of a composite score under the assumptions of the measurement models used in classic test theory (Lord and Novick, 1968, pp. 88-90) where reliability is defined as the proportion of observed score variation represented by true score (as opposed to error) variation or, equivalently, the correlation between true and observed or fallible scores.

The Kuder-Richardson coefficient also has an internal consistency interpretation of particular interest here in that it represents the proportion of test variance due to all common factors present in the test (Cronbach, 1951, pp. 319-321), assuming the items are "essentially tau equivalent" or have equal units of measurement (Lord and Novick, 1968, pp. 88-90). The items comprising each of the composite measurement discussed were selected on the basis of results from factor analyses reported in other studies typically conducted among other disparate U.S. samples, which indicated that these items loaded heavily on a common underlying factor. This parallels a fairly widespread practice in cross-national marketing research wherein one develops a measurement instrument in one language and

setting and then subsequently applies it in one or more other different locations after translation into the appropriate additional languages. Under such circumstances, it is well to ask, prior to carrying out any statistical analyses of tests of hypotheses using the composite scores, whether or not the alternative linguistic versions of the instrument have the same internal consistency properties. A comparison of the Kuder-Richardson coefficients across samples can serve as a preliminary diagnostic check on this issue and can be implemented with relative ease in the context of generating the types of simple summary statistics typically required in preliminary analyses of survey data. It bears noting that Warren, White and Fuller (1974) have proposed an estimation procedure for a linear regression model with errors in the variables that makes use of Cronbach  $\alpha$  or split-half reliability coefficients to correct for attenuation produced by measurement error where the items comprising the composite scores conform to the assumptions of classic test theory. The confirmatory factor analytic and structural equation methods developed by Joreskog (1970, 1971) and applied to marketing research problems by Bagozzi (1980) offer a more general method of dealing with this class of problems which allows reliability and validity assessment to be integrated with the estimation of parameters for larger models consisting of imperfectly measured variables. In particular, the Joreskog procedure can be used to conduct a formal test of the equivalence of the factor structure of items across different samples. Given that the present interest in the data at hand is in using them to explore the overall need for detecting cross-national variations in the quality of measurements rather than in investigating substantive issues relating to the use of these particular measurements, we confine our analysis of composite scores here to a comparison of Kuder-Richardson coefficients.

## RESULTS

### Background Characteristics

Table 2 shows the coefficients of agreement, by sample, for the set of 10 demographic and other more or less "invariant" household characteristics typical of the kinds of background variables routinely measured in consumer surveys. Ninety per cent confidence intervals were computed for each coefficient using the approximate large sample variance for  $\kappa$  given by Bishop, Fienberg, and Holland (1975, p. 396) and are shown in parentheses in Table 2. An inspection of Table 2 reveals that the  $\kappa$  values are generally quite high, indicating strong and consistent reliability across variables and samples. All fifty coefficients are significantly different from zero ( $p < .01$ ), more than half (28) exceed .9, and 84 per cent (42/50) exceed .8. To identify between-sample differences in the levels of agreement, we examine each row of Table 2 for non-overlapping confidence intervals. For three of the ten variables, a single pairwise difference occurred and all three of these differences involve the Quebec sample. For "husband's age" and "type of dwelling", the Quebec coefficients appear significantly lower than the relevant Chicago statistics while in the case of "number of appliances", the London/Glasgow-Quebec contrast indicates different levels of agreement. Rank-ordering the  $\kappa$  coefficients by city for each variable, one finds the lowest levels of agreement occur in the Quebec sample for five of the ten background characteristics. However, when we examine the extent of association in the rankings of the  $\kappa$  coefficients for five samples across the ten variables, we find no more consistency than would be expected by chance. The value of Kendall's coefficient of concordance (Siegel, 1956, pp. 229-238) which measures the degree of correspondence among several rankings of the same set of objects (here, ten rankings, one

per variable, of five  $\kappa$  coefficients, one per sample per variable), was found to be +.124 which is not significantly different from zero ( $\chi^2 = 4.94$ ,  $df = 4$ ,  $p > .20$ ). Thus, there is no firm evidence of persistent differences in reliability among the samples for all the background variables investigated.

-----

INSERT TABLE 2 HERE

-----

Examining the levels of husband-wife congruence for different variables, we observe that the  $\kappa$  values for four variables (husband's education, wife's education, appliance ownership, and income) tend to be somewhat lower as compared to those for the other six characteristics. The eight  $\kappa$  coefficients shown in Table 2 less than .8 were all associated with the aforementioned four variables. These four characteristics, especially appliance ownership and income, are ones about which one spouse is likely to be better informed than the other due to role specialization and the division of labor within a family. It is interesting to note that husbands and wives in all five samples appear somewhat less likely to agree in reporting each other's educational backgrounds than each other's ages. Perhaps in the normal course of events a spouse is more frequently asked or reminded of his/her partner's age as compared to his/her education.

Overall, the results in Table 2 indicate a high degree of reliability in all five samples for the set of ten background variables investigated here. Relatively few differences in classifying households on the majority of these variables would occur by relying on one spouse rather than the

other to supply the information required to do so. Table A1 presented in the Appendix displays the percentages of households, by sample, for which the husband and wife gave identical reports of the various background characteristics discussed above. The percentage of households wherein husbands and wives differed in their reports was 5-10 per cent in the case of standard demographic variable (length of marriage, number of children, age of spouse, wife's employment status, and type of dwelling unit) and 10-15 per cent for the variables requiring greater familiarity or specialized information (spouse's education, appliance ownership, and income).

These results compare favorably with findings of other similar U.S. studies reported in the literature. To illustrate, whereas Ferber (1955) and Haberman and Elinson (1967) found 71 per cent and 60 per cent, respectively, of responding couples reported income in the same scale category, here we observed agreement levels from 84 to 93 per cent for income (see Table A1). Schreiber (1975-76) investigated the reliability of age and education using data from test-retest studies conducted over two year intervals in Britain and the U.S.. For age, 91 per cent of the U.S. sample and 98 per cent of the British sample gave consistent reports in both waves of measurement. However, for education, the figure was only 74 per cent in the U.S. sample. Again, these results are quite similar to the present findings (see Table A1) and suggest that spousal agreement studies may yield reliability estimates of background variables comparable to those obtained from test-retest designs. The considerable variability in the reliability estimates of demographic variables reported in the literature serves as a reminder that the quality of measurements of even seemingly

"invariant" characteristics of individuals is not something that can be taken for granted. Indeed, the level of reliability attained for background characteristics may serve as a benchmark or standard to which we may aspire in efforts to measure more complex and elusive attitudinal and behavioral constructs of the kind we consider below.

#### Husband-Wife Task and Decision Involvement

Husband-wife congruence in reporting involvement in five family task/decision areas was assessed in the same manner described above with reference to background characteristics. Recall that the involvement ratings were a trichotomous scale, distinguishing between whether responsibility for a decision or task was "mainly husband" vs. joint or shared vs. "mainly wife". Table 3 below shows the coefficients of agreement for husbands' and wives' reports of involvement for each decision/task area and sample.

-----  
INSERT TABLE 3 HERE  
-----

It is immediately apparent from Table 3 that there is considerably less agreement between husbands and wives about task/decision involvement than was observed for background characteristics. Only half the twenty-five coefficients in Table 3 exceed .5 and a fifth are less than .4. The lower bounds of the 90 per cent confidence intervals exceed zero for all 25 coefficients shown in Table 3 and we conclude that husband and wife agreement regarding involvement in these tasks/decisions exceeds levels expected by chance but, in general, is rather moderate.

A handful of between-sample differences are discernible from non-overlapping confidence levels in Table 3, but they are not concentrated around any particular sample or task/decision area. For "savings and investment", agreement in the Quebec sample is lower than for the Chicago sample, while for "travel reservations", Chicago is lower than London/Glasgow. In the case of two areas, "dining out" and "automobile shopping", no significant pairwise differences are apparent. However for "supermarket shopping", agreement in the Brussels sample appears lower than that in both the Chicago and Quebec samples. The absence of any overall pattern of systematic between-sample differences is further reinforced by the fact that the association in the rankings of the coefficients for the five samples across the five task/decision areas, as measured by Kendall's coefficient of concordance, is only .144, which is not significantly different from zero ( $\chi^2 = 2.88$ ,  $df = 4$ ,  $p > .5$ ). Finally, comparing the levels of agreement for the five tasks/decisions, we detect no tendency for any particular variables to be consistently high or low in all five samples.

In contrast to the background characteristics discussed previously, reliance on one spouse for information regarding task/decision involvement would make a very substantial difference in how households would be classified with respect to these variables. The proportion of cases where husbands and wives gave identical responses in reporting involvement varied from 62 to 89 per cent with most cases falling in the 70-75 per cent range. Detailed results by task/decision area and sample are presented in the Appendix, Table A2. This level of agreement compares favorably with the 30-80 per cent range obtained in other similar studies reviewed by

Davis (1976). Given the considerable incidence of disagreement observed here, it is of interest to know whether the differences between husbands' and wives' ratings tend to be random or biased in either the direction of "modesty" (underestimating one's own involvement and overestimating that of one's spouse's) or "vanity" (overestimating one's own involvement and underestimating that of one's spouse)(Davis and Rigaux, 1974, p. 58). For each task/decision and sample, we compared the relative frequency of modesty versus vanity disagreements and found that in about two-thirds (17/25) of the comparisons, the latter was more prevalent than the former. This "vanity" bias occurred in all five samples but was somewhat more pronounced for London/Glasgow respondents. The tendency toward seeming self-aggrandizement observed here may be a function of the task/decision areas investigated inasmuch as Davis and Rigaux (1974, p. 59) note that conflicting results have been reported in the literature as to the prevalence of modesty or vanity biases and conclude that "one cannot generalize about the likelihood of encountering either of these perceptual biases."

In light of the results presented above which indicated only a moderate degree of husband-wife agreement plus the presence of some response bias, the question arises as to whether these ratings constitute meaningful reports about who is involved in the specific task/decision areas to which the original questionnaire items referred, or instead are more of a reflection of socially acceptable stereotypes respondents hold about who should be involved in these matters. To the extent that repeated application of the same rating scale within respondents elicits response biases and other sources of irrelevant method variance common to the particular instrument employed, then differences in involvement across task/decision areas will be obscured and the ability to discriminate among

the areas on the basis of the observed ratings will be adversely affected. In order to assess this matter we utilize Campbell and Fiske's (1959) multitrait-multimethod approach to discriminant validation which Davis (1971) has previously applied to husband-wife ratings data. Briefly, we treat the set of ten husbands' and wives' ratings as two different methods of measuring each of five traits (i.e., task/decision areas). For each sample, we compute all possible pairwise correlations among the ten ratings using Goodman and Kruskal's (1954) gamma as a summary measure of association. The set of associations, arranged in Campbell and Fiske's multitrait-multimethod matrix format (one per sample), are presented as Tables A3-A7 in the Appendix. The basic principle underlying Campbell and Fiske's procedures is that different methods of measuring the same trait should correlate more strongly with each other than they do with measures of other, purportedly different traits. Campbell and Fiske (1959) suggest two tests for discriminant ability. Translated into the present context, the requirements are:

1. Husbands' (wives') ratings for a particular task/decision should be more strongly associated with wives' (husbands') ratings for the same task/decision than with wives' (husbands') ratings for a different task/decision.
2. Husbands' (wives') ratings for a particular task/decision should be more strongly associated with wives' (husbands') ratings for the same task/decision than with the husbands' (wives') ratings for a different task/decision.

Testing for consistency with these criteria involves a large number of comparisons between pairs of gamma coefficients -- forty comparisons per sample for the first criterion and another set of forty comparisons per sample for the second criterion. With virtual unanimity, the results obtained for both requirements were confirmatory for all five samples.<sup>2/</sup>

The "vanity" response bias discussed above appears not to have been sufficiently strong to overwhelm the discriminant quality of the measures.<sup>3/</sup>

To sum up then, we find husbands' and wives' involvement ratings exhibit only moderate levels of convergence, which is subject to some between-sample variability but the measures do possess good discriminant ability despite indications of the presence of "vanity" response bias.

### Life-Style Variables

Presented below are Kuder-Richardson coefficients for each of six life-style variables and for each of the five samples. Table 4 contains the coefficients for female respondents and Table 5 for male respondents. Also presented in these tables are 90 per cent confidence intervals for each coefficient, computed using the method proposed by Feldt (1965).

-----  
INSERT TABLES 4 AND 5 HERE  
-----

It is readily apparent from an examination of Tables 3 and 4 that the Kuder-Richardson coefficients vary markedly across both samples and variables. The median value of the 30 coefficients for the female respondents is .504 (range: .00-.728) and .533 (range: .155-.844) for male respondents. Inspecting the 90 per cent confidence intervals, one would conclude that one of female coefficients (task allocation in the Quebec sample) and three of the male coefficients (orderliness in the London/Glasgow sample and both task allocation and orderliness in the Quebec sample) are not significantly different from zero. A cell-by-cell comparison of Tables 4 and 5 indicates no simple overall pattern of sex differences in reliability; for about half (16) the 30 coefficients, the values are greater for male respondents than for females while the reverse is true for the other half.

Examining each row of Tables 4 and 5 for nonoverlapping confidence intervals, we find one or more pairs of between-sample differences in the Kuder-Richardson coefficients for four of the six variables in the case of female respondents but for only two of six in the case of male respondents. For female respondents, there is no consistent pattern in the rank order of the samples with respect to reliability across all six variables; Kendall's coefficient of concordance for Table 4 was found to have a nonsignificant value of .067 ( $\chi^2 = 1.333$ ,  $df = 4$ ,  $p > .8$ ). However, such a pattern does manifest itself in the male data in Table 5 where there is some similarity across these six variables in the rank order of the five samples with respect to reliability as indicated by a value of .417 for the Kendall concordance coefficient ( $\chi^2 = 8.33$ ,  $df = 4$ ,  $.05 < p < .10$ ). The highest reliability coefficients tend to be found in the male Paris sample while Brussels and Quebec tend to be low with Chicago and London/Glasgow occupying a middling position. Note that this pattern of variation in the reliability coefficients is not related in any obvious way to language differences. The two English-speaking samples were bracketed by the three French-speaking samples in overall rank order of reliability for the female respondents.

How do these estimates compare with the reliability reported for other life-style measurement instruments? Contrasting the handful of reliability results reported in the literature (Wells, 1975, pp. 202-203) is hazardous because of differences in the number of items forming the composite scores and because estimates reported previously appear to consist entirely of test-retest coefficients whose strict interpretation as a measure of reliability may be equivocal for reasons discussed in Silk (1977). Subject to those qualifications, it can be mentioned that Pessemier and Bruno (1971, p. 392) reported test-retest correlations (six month time interval) for twelve multi-item psychographic dimensions which ranged from .64 to .9

with a median of .78, values which are somewhat higher than the coefficients show in Tables 4 and 5.

## SUMMARY AND DISCUSSION

The motivation for this paper was the concern that linguistic and conceptual non-equivalencies in questionnaire instruments used in cross national surveys could operate to produce differences in the reliabilities of measurements. The latter condition would pose a threat to the validity of conclusions drawn about similarities or differences in markets on the basis of comparisons among measures which varied in extent of their fallibility. The results reported here provide confirmation, however unwelcome, for the suspicion that significant between-sample reliability differentials can arise in cross national surveys employing instruments developed through the use of simple but commonplace back translation procedures which attempt to achieve linguistic and conceptual equivalence. Although between-sample differences were uncovered, there was no systematic tendency for any particular national sample or language group to exhibit high or low reliability consistently across several different types of measurements or across several different variables of the same type. However, the incidence of reliability differentials did seem to vary according to the type of variable, being less likely to occur for "hard" variables like demographic characteristics and more likely to occur for "soft" variables such as task/decision involvement and life style/psychographic factors. Thus it would appear that the attainment of linguistic and conceptual equivalence in cross-national surveys is more difficult for attitudinal and perceptual variables than for demographic and other background

variables. Such a result is especially discouraging in light of the fact that the demographic variables were measured here by single items while the life style measures were derived from multiple items. Although direct comparisons between the agreement and internal consistency coefficients used as indices of reliability for these two types of variables are problematic, it may be that the kinds of error factors operating here are such that they exhibit heteroskedastic behavior--i.e., as the mean level of reliability decreases, its variability increases.

We noted at the outset of this paper that measure unreliability attenuates the precision of estimators and reduces the power of statistical tests. Given that, the principal implication of these findings is that in addition to making efforts to control and minimize non-equivalencies, more attention should be directed toward diagnosing their effects on reliability and making adjustments for unreliability in carrying out statistical analysis involving cross-national comparisons. Such a strategy of detection and correction for unreliability would require that reliability estimates be made available routinely for cross-national surveys, just as the reporting of interview verification and sampling errors has become an accepted part of research practice. For multiple item scales, internal consistency reliability statistics can be derived from the same study. However, for single item measures, which tend to be the most common, it would be necessary to obtain reliability estimates in other ways, such as by interviewing both husbands and wives and/or by conducting special test-retest studies. While the difficulty and expense of such undertakings are not pleasant to contemplate, neither is the prospect of drawing invalid conclusions from cross-national studies by ignoring this issue. Few marketing researchers would care to practice their trade without the information that comes from pre-tests, interview verifications, and sampling error

calculations. Reliability assessment deserves to occupy a similar status in the quality control process for cross-national marketing research.

Having called attention to the problem, we do need to emphasize that means of coping with it are available. More than a decade ago, Brown (1969) urged that non-sampling errors be taken seriously in using marketing research results and proposed a method, "credence analysis", for making subjective assessments of their accuracy. Since that time, important developments have taken place in econometrics and psychometrics (Goldberger, 1971 and Griliches, 1974) leading to the availability of a considerable body of formal statistical methods, including software packages, that take explicit account of errors-in-variables problems. A recent book by Bagozzi (1980) provides an excellent review of much relevant material with applications in marketing research. All of this points to escalating the complexity of carrying out cross-national marketing research, but then who ever said that multinational marketing was going to be easy?

# FOOTNOTES

1. Establishing convergence between different measures of the same trait is, of course, prerequisite to assessing their discriminant ability. In the earlier discussion we noted that the coefficients of agreement between husbands' and wives' involvement ratings were all statistically significant and consistent with those results, the corresponding gamma statistics measuring the convergent association (as opposed to agreement in the sense of identical ratings) were also all found to be substantial and statistically significant. See Tables A3-A7 in the appendix.
2. A single comparison failed to comply with the second requirement. This occurred in the Quebec sample where the convergent association between husbands' and wives' ratings for automobile shopping (.512) turned out to be smaller than the association between wives' ratings of automobile shopping and savings/investment (.583).
3. Campbell and Fiske (1959) suggest a third criteria for discriminant ability, namely that the same pattern of trait interrelationships should hold in all of the heterotrait triangles of both the mono-method and heteromethod blocks. Here there are four heterotrait triangles, each containing ten intercorrelations. Kendall's coefficient of concordance was computed to assess the degree of consistency in the rank order of the heterotrait associations across the four triangles. The results obtained for all five samples were consistent with the requirements for discrimination and are summarized below.

| Sample         | Coeff. of Concordance | $\chi^2$<br>(df=9) | P      |
|----------------|-----------------------|--------------------|--------|
| Chicago        | .772                  | 27.81              | < .01  |
| London/Glasgow | .745                  | 26.84              | < .01  |
| Paris          | .892                  | 32.13              | < .001 |
| Brussels       | .872                  | 31.42              | < .001 |
| Quebec         | .573                  | 20.62              | < .02  |

TABLE 1  
MEASUREMENT

| TYPE                                     | VARIABLES                                                                                                                                                                                     | SCALES                                              |
|------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| Background Characteristics               | Length of Marriage<br>Number of Children<br>Husband's Age<br>Wife's Age<br>Wife's Employment<br>Husband's Education<br>Wife's Education<br>Type of Dwelling<br>Number of Appliances<br>Income | Categorical<br><br>Single Item per Variable         |
| Involvement in Household Tasks/Decisions | Savings and Investments<br>Travel Reservations<br>Dining Out<br>Shopping for an Automobile<br>Shopping at Supermarkets                                                                        | Three Point Ratings<br><br>Single Item per Variable |
| Life Cycle                               | Male Dominance<br>Task Allocation<br>Female Role Perception<br>Orderliness<br>Anxiety and Control<br>Traditionalism                                                                           | Composite Scores<br><br>Multiple Items per Variable |

TABLE 2

COEFFICIENTS OF AGREEMENT (K) BETWEEN HUSBANDS' AND WIVES' REPORTS  
OF SELECTED BACKGROUND CHARACTERISTICS BY SAMPLE\*

| Background Characteristic** | Sample              |                     |                     |                     |                     |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                             | Chicago             | London/<br>Glasgow  | Paris               | Brussels            | Quebec              |
| Length of Marriage (5)      | .967<br>(.940~.994) | .957<br>(.916~.998) | .976<br>(.948~1.0)  | .951<br>(.896~1.0)  | .944<br>(.901~.987) |
| Number of Children (5)      | .953<br>(.922~.984) | .985<br>(.960~1.0)  | .926<br>(.878~.974) | .980<br>(.947~1.0)  | .951<br>(.916~.984) |
| Husband's Age (4)           | .974<br>(.949~.999) | .955<br>(.912~.998) | .975<br>(.944~1.0)  | .956<br>(.906~1.0)  | .871<br>(.802~.939) |
| Wife's Age (4)              | .973<br>(.947~.999) | .968<br>(.931~1.0)  | .960<br>(.923~.997) | .928<br>(.861~.995) | .931<br>(.889~.973) |
| Wife's Employment (2)       | .937<br>(.891~.983) | .979<br>(.945~1.0)  | .981<br>(.951~1.0)  | 1.0                 | 1.0                 |
| Husband's Education (4)     | .796<br>(.724~.868) | .797<br>(.712~.882) | .816<br>(.738~.894) | .874<br>(.792~.956) | .871<br>(.794~.948) |
| Wife's Education (4)        | .823<br>(.747~.899) | .783<br>(.728~.838) | .815<br>(.732~.898) | .745<br>(.624~.866) | .745<br>(.638~.851) |
| Type of Dwelling (2)        | .985<br>(.961~1.0)  | .874<br>(.772~.976) | .957<br>(.907~1.0)  | 1.0                 | .851<br>(.761~.939) |
| Number of Appliances (6)    | .813<br>(.756~.870) | .858<br>(.788~.928) | .794<br>(.718~.870) | .842<br>(.754~.930) | .663<br>(.558~.768) |
| Income (6)                  | .852<br>(.797~.907) | .833<br>(.758~.908) | .888<br>(.827~.949) | .901<br>(.830~.972) | .792<br>(.709~.875) |

\* A 90% confidence interval is shown in parentheses for each K value.

\*\* The figure in parentheses next to each background characteristic represents the number of response categories appearing in the questionnaire item.

TABLE 3

COEFFICIENTS OF AGREEMENT (K) BETWEEN HUSBANDS' AND WIVES'

RATINGS OF INVOLVEMENT FOR SELECTED

HOUSEHOLD TASKS AND DECISIONS BY SAMPLE\*

| Task/Decision              | Sample              |                     |                     |                     |                     |
|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                            | Chicago             | London/<br>Glasgow  | Paris               | Brussels            | Quebec              |
| Savings & Investments      | .623<br>(.528~.718) | .605<br>(.473~.737) | .541<br>(.422~.664) | .492<br>(.336~.648) | .312<br>(.148~.477) |
| Travel Reservations        | .292<br>(.172~.412) | .639<br>(.502~.776) | .484<br>(.364~.604) | .537<br>(.392~.682) | .397<br>(.254~.540) |
| Dining Out                 | .496<br>(.396~.596) | .408<br>(.253~.563) | .580<br>(.448~.712) | .624<br>(.486~.762) | .374<br>(.208~.540) |
| Shopping for an Automobile | .420<br>(.295~.545) | .513<br>(.353~.673) | .422<br>(.277~.567) | .468<br>(.300~.636) | .252<br>(.073~.431) |
| Shopping at Supermarket    | .715<br>(.611~.819) | .575<br>(.407~.743) | .652<br>(.534~.770) | .437<br>(.276~.598) | .774<br>(.677~.871) |

\* A 90% confidence interval is shown in parentheses  
for each K value.

TABLE 4  
RELIABILITY COEFFICIENTS (KUDER-RICHARDSON) FOR SELECTED COMPOSITE  
LIFE-STYLE MEASURES BY SAMPLE: FEMALE RESPONDENTS\*

| Measure**                  | Sample              |                     |                     |                     |                     |
|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                            | Chicago             | London/<br>Glasgow  | Paris               | Brussels            | Quebec              |
| Male Dominance (4)         | .728<br>(.660~.782) | .629<br>(.517~.725) | .592<br>(.470~.690) | .637<br>(.510~.742) | .313<br>(.093~.499) |
| Task Allocation (5)        | .407<br>(.265~.526) | .443<br>(.287~.597) | .567<br>(.446~.667) | .379<br>(.168~.553) | (-.083)<br>(0~.260) |
| Female Role Perception (4) | .504<br>(.380~.603) | .615<br>(.492~.715) | .587<br>(.463~.686) | .561<br>(.408~.688) | .549<br>(.405~.666) |
| Orderliness (4)            | .411<br>(.264~.529) | .336<br>(.124~.509) | .439<br>(.270~.573) | .442<br>(.247~.604) | .464<br>(.292~.600) |
| Anxiety and Control (4)    | .338<br>(.172~.470) | .467<br>(.297~.606) | .432<br>(.262~.568) | .633<br>(.505~.740) | .415<br>(.228~.566) |
| Traditionalism (7)         | .583<br>(.487~.662) | .607<br>(.497~.701) | .582<br>(.469~.682) | .298<br>(.073~.481) | .674<br>(.576~.750) |

\* A 90% confidence interval is shown in parenthesis for each Kuder-Richardson coefficient.

\*\* The figure in parentheses next to each measure is the number of items comprising the composite scale.

TABLE 5

RELIABILITY COEFFICIENTS (KUDER-RICHARDSON) FOR SELECTED COMPOSITE  
LIFE-STYLE MEASURES BY SAMPLE: MALE RESPONDENTS\*

| Measure**                  | Sample              |                     |                     |                     |                     |
|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                            | Chicago             | London/<br>Glasgow  | Paris               | Brussels            | Quebec              |
| Male Dominance (4)         | .622<br>(.527~.697) | .672<br>(.573~.757) | .684<br>(.589~.760) | .592<br>(.449~.710) | .504<br>(.345~.633) |
| Task Allocation (5)        | .478<br>(.348~.583) | .542<br>(.414~.652) | .577<br>(.459~.674) | .425<br>(.236~.586) | .155<br>(0~.375)    |
| Female Role Perception (4) | .579<br>(.474~.664) | .563<br>(.432~.677) | .538<br>(.400~.649) | .533<br>(.369~.668) | .520<br>(.366~.645) |
| Orderliness (4)            | .844<br>(.805~.875) | .243<br>(.001~.450) | .530<br>(.399~.643) | .388<br>(.174~.565) | .155<br>(0~.374)    |
| Anxiety and Control (4)    | .412<br>(.265~.530) | .450<br>(.273~.593) | .536<br>(.397~.648) | .522<br>(.355~.661) | .506<br>(.348~.635) |
| Traditionalism (7)         | .561<br>(.460~.645) | .518<br>(.384~.634) | .616<br>(.508~.708) | .353<br>(.146~.521) | .568<br>(.439~.676) |

\* A 90% confidence interval is shown in parenthesis for each Kuder-Richardson coefficient.

\*\* The figure in parentheses next to each measure is the number of items comprising the composite scale.



## APPENDIX

Footnote to Tables A3-A7:

The significance levels noted in these tables refer to  $\chi^2$  tests of the null hypothesis of no association in the 3x3 contingency tables from which the corresponding gamma statistics reported there were calculated.



Table A1

PERCENT OF HOUSEHOLDS FOR WHICH HUSBAND AND WIFE GAVE IDENTICAL  
REPORTS OF SELECTED BACKGROUND CHARACTERISTICS, BY SAMPLE

| Background<br>Characteristic | Sample  |                    |       |          |        |
|------------------------------|---------|--------------------|-------|----------|--------|
|                              | Chicago | London/<br>Glasgow | Paris | Brussels | Quebec |
| Length of Marriage           | 97.4    | 96.9               | 98.1  | 97.0     | 96.6   |
| Number of Children           | 96.3    | 98.8               | 94.4  | 98.4     | 96.7   |
| Husband's Age                | 98.1    | 96.8               | 98.1  | 97.3     | 91.0   |
| Wife's Age                   | 98.1    | 97.9               | 97.2  | 95.9     | 95.6   |
| Wife's Employment            | 96.9    | 99.0               | 99.1  | 100.1    | 100.0  |
| Husband's Education          | 88.2    | 85.9               | 86.6  | 91.9     | 90.6   |
| Wife's Education             | 91.7    | 96.3               | 87.4  | 86.1     | 85.4   |
| Type of Dwelling             | 99.4    | 95.8               | 98.1  | 100.0    | 93.3   |
| Number of Appliances         | 84.3    | 89.5               | 84.8  | 86.5     | 76.9   |
| Income                       | 89.2    | 87.2               | 90.9  | 92.6     | 83.7   |

Table A2

PERCENT OF HOUSEHOLDS FOR WHICH HUSBAND AND WIFE GAVE IDENTICAL RATINGS  
OF INVOLVEMENT IN SELECTED TASKS AND DECISIONS, BY SAMPLE

| Task/Decision              | Sample  |                    |       |          |        |
|----------------------------|---------|--------------------|-------|----------|--------|
|                            | Chicago | London/<br>Glasgow | Paris | Brussels | Quebec |
| Savings & Investment       | 79.6    | 78.6               | 74.8  | 72.6     | 67.4   |
| Travel Reservations        | 70.7    | 79.2               | 66.0  | 71.9     | 64.0   |
| Dining Out                 | 71.3    | 69.9               | 77.8  | 78.9     | 71.6   |
| Shopping for an Automobile | 72.5    | 74.6               | 70.0  | 72.3     | 61.5   |
| Shopping at Supermarket    | 89.4    | 83.3               | 82.4  | 70.8     | 86.5   |

Table A 3

## ASSOCIATIONS (GAMMAS) AMONG HUSBANDS' AND WIVES' INVOLVEMENT RATINGS: CHICAGO SAMPLE

| Respondent \ Task or Decision* | Husband           |                    |                   |                   |                   | Wife              |                    |       |                    |   |
|--------------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|-------------------|--------------------|-------|--------------------|---|
|                                | F                 | T                  | D                 | C                 | M                 | F                 | T                  | D     | C                  | M |
| Husband                        | -                 |                    |                   |                   |                   |                   |                    |       |                    |   |
|                                | .452 <sup>a</sup> | -                  |                   |                   |                   |                   |                    |       |                    |   |
|                                | .063              | .207               | -                 |                   |                   |                   |                    |       |                    |   |
|                                | .277              | .253 <sup>b</sup>  | .208 <sup>b</sup> | -                 |                   |                   |                    |       |                    |   |
|                                | -.098             | -.167              | .063              | -.441             | -                 |                   |                    |       |                    |   |
| Wife                           | .938 <sup>a</sup> | .400 <sup>a</sup>  | .240              | .259              | -.172             | -                 |                    |       |                    |   |
|                                | .402 <sup>a</sup> | .820 <sup>a</sup>  | .157              | .070              | -.278             | .384 <sup>a</sup> | -                  |       |                    |   |
|                                | .008              | -.062              | .721 <sup>a</sup> | .041              | .012              | .099              | .201               | -     |                    |   |
|                                | -.002             | .331 <sup>c</sup>  | .032              | .764 <sup>a</sup> | -.154             | .146              | .226 <sup>b</sup>  | -.027 | -                  |   |
|                                | .010              | -.318 <sup>b</sup> | .154              | -.401             | .956 <sup>a</sup> | -.127             | -.320 <sup>c</sup> | -.274 | -.556 <sup>b</sup> | - |

<sup>a</sup>  $p < .01$ <sup>b</sup>  $p < .05$ <sup>c</sup>  $p < .10$ 

\*F = Savings and Investment

T = Travel Reservations

D = Dining Out

C = Shopping for an Automobile

M = Shopping at Supermarket

Table A 4

## ASSOCIATIONS (GAMMAS) AMONG HUSBANDS' AND WIVES' INVOLVEMENT RATINGS: LONDON/GLASGOW SAMPLE

| Respondent \ Task or Decision* | Husband                     |                             |                             |                             |                              | Wife               |      |                   |      |   |
|--------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|------------------------------|--------------------|------|-------------------|------|---|
|                                | F                           | T                           | D                           | C                           | M                            | F                  | T    | D                 | C    | M |
| Husband                        | -                           |                             |                             |                             |                              |                    |      |                   |      |   |
|                                | .354                        |                             |                             |                             |                              |                    |      |                   |      |   |
|                                | -.083                       | .342 <sup>a</sup>           | -                           |                             |                              |                    |      |                   |      |   |
|                                | .048                        | .074                        | .580 <sup>b</sup>           | -                           |                              |                    |      |                   |      |   |
|                                | .010                        | .205 <sup>b</sup>           | .127                        | -.342                       | -                            |                    |      |                   |      |   |
| Wife                           | <del>.874<sup>a</sup></del> | <del>.669<sup>a</sup></del> | <del>-.037</del>            | <del>.008</del>             | <del>-.013<sup>b</sup></del> | -                  |      |                   |      |   |
|                                | <del>.333<sup>c</sup></del> | <del>.935<sup>a</sup></del> | <del>.198<sup>a</sup></del> | <del>.104</del>             | <del>.380</del>              | .707 <sup>a</sup>  | -    |                   |      |   |
|                                | <del>.251</del>             | <del>.421<sup>a</sup></del> | <del>.820<sup>a</sup></del> | <del>.270</del>             | <del>-.058<sup>c</sup></del> | .379 <sup>b</sup>  | .394 | -                 |      |   |
|                                | <del>.175</del>             | <del>.283</del>             | <del>.571<sup>b</sup></del> | <del>.885<sup>a</sup></del> | <del>-.106</del>             | -.029              | .296 | .574 <sup>a</sup> | -    |   |
|                                | <del>-.382</del>            | <del>.051</del>             | <del>.139</del>             | <del>.143</del>             | <del>.891<sup>a</sup></del>  | -.275 <sup>b</sup> | .249 | -.322             | .151 | - |

<sup>a</sup>  $p < .01$ <sup>b</sup>  $p < .05$ <sup>c</sup>  $p < .10$ 

\*F = Savings and Investment

T = Travel Reservations

D = Dining Out

C = Shopping for an Automobile

M = Shopping at Supermarket

Table A 5

## ASSOCIATIONS (GAMMAS) AMONG HUSBANDS' AND WIVES' INVOLVEMENT RATINGS: PARIS SAMPLE

| Respondent \ Task or Decision* | Husband |                   |                   |                   |                   | Wife               |                   |       |                   |       |
|--------------------------------|---------|-------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------|-------------------|-------|
|                                | F       | T                 | D                 | C                 | M                 | F                  | T                 | D     | C                 | M     |
| Husband                        | F       | -                 |                   |                   |                   |                    |                   |       |                   |       |
|                                | T       | .205              | -                 |                   |                   |                    |                   |       |                   |       |
|                                | D       | .325              | .067              | -                 |                   |                    |                   |       |                   |       |
|                                | C       | .459 <sup>a</sup> | .123              | .339 <sup>a</sup> | -                 |                    |                   |       |                   |       |
|                                | M       | -.005             | .158              | -.048             | -.006             | -                  |                   |       |                   |       |
| Wife                           | F       | .857 <sup>a</sup> | .197              | .327 <sup>c</sup> | .352              | -.198 <sup>b</sup> | -                 |       |                   |       |
|                                | T       | .213              | .832 <sup>a</sup> | .111              | .176 <sup>c</sup> | -.055              | .335 <sup>b</sup> | -     |                   |       |
|                                | D       | .221              | .086              | .935 <sup>a</sup> | .400 <sup>b</sup> | -.120              | .276              | .010  | -                 |       |
|                                | C       | .574 <sup>b</sup> | -.019             | .145              | .753 <sup>a</sup> | .095               | .520 <sup>b</sup> | .157  | .194 <sup>b</sup> | -     |
|                                | M       | -.160             | .180              | -.023             | -.057             | .915 <sup>a</sup>  | -.229             | -.052 | -.206             | -.152 |

<sup>a</sup>  $p < .01$ 

\*F = Savings and Investment

<sup>b</sup>  $p < .05$ 

T = Travel Reservations

<sup>c</sup>  $p < .10$ 

D = Dining Out

C = Shopping for an Automobile

M = Shopping at Supermarket

Table A 6  
ASSOCIATIONS (GAMMAS) AMONG HUSBANDS' AND WIVES' INVOLVEMENT RATINGS: BRUSSELS SAMPLE

| Respondent \ Task or Decision* | Husband |                             |                             |                             |                             | Wife                        |                    |       |      |                    |
|--------------------------------|---------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|--------------------|-------|------|--------------------|
|                                | F       | T                           | D                           | C                           | M                           | F                           | T                  | D     | C    | M                  |
| Husband                        | F       | -                           |                             |                             |                             |                             |                    |       |      |                    |
|                                | T       | .531 <sup>b</sup>           | -                           |                             |                             |                             |                    |       |      |                    |
|                                | D       | -.314                       | -.122                       | -                           |                             |                             |                    |       |      |                    |
|                                | C       | .479 <sup>a</sup>           | .180                        | -.251                       | -                           |                             |                    |       |      |                    |
|                                | M       | -.156                       | -.121                       | .539 <sup>c</sup>           | -.467                       | -                           |                    |       |      |                    |
| Wife                           | F       | <del>.864<sup>a</sup></del> | <del>.485<sup>a</sup></del> | <del>-.179</del>            | <del>.524<sup>a</sup></del> | <del>-.249</del>            | -                  |       |      |                    |
|                                | T       | <del>.507<sup>b</sup></del> | <del>.888<sup>a</sup></del> | <del>-.026</del>            | <del>.078</del>             | <del>-.014</del>            | .504 <sup>b</sup>  | -     |      |                    |
|                                | D       | <del>-.223</del>            | <del>-.370</del>            | <del>.899<sup>a</sup></del> | <del>-.229</del>            | <del>.493</del>             | -.420 <sup>c</sup> | -.264 | -    |                    |
|                                | C       | <del>.731<sup>b</sup></del> | <del>.176</del>             | <del>-.039</del>            | <del>.847<sup>a</sup></del> | <del>-.489</del>            | .551 <sup>b</sup>  | .125  | .116 | -                  |
|                                | M       | <del>-.111</del>            | <del>-.182</del>            | <del>.218</del>             | <del>-.465</del>            | <del>.820<sup>a</sup></del> | -.421              | -.282 | .266 | -.328 <sup>c</sup> |

<sup>a</sup>  $p < .01$

<sup>b</sup>  $p < .05$

<sup>c</sup>  $p < .10$

\*F = Savings and Investment

T = Travel Reservations

D = Dining Out

C = Shopping for an Automobile

M = Shopping at Supermarket

Table A 7

## ASSOCIATIONS (GAMMAS) AMONG HUSBANDS' AND WIVES' INVOLVEMENT RATINGS: QUEBEC SAMPLE

| Respondent \ Task or Decision* | Husband |                             |                              |                             |                             | Wife                        |                   |                   |      |       |
|--------------------------------|---------|-----------------------------|------------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------|-------------------|------|-------|
|                                | F       | T                           | D                            | C                           | M                           | F                           | T                 | D                 | C    | M     |
| Husband                        | F       | -                           |                              |                             |                             |                             |                   |                   |      |       |
|                                | T       | .498 <sup>a</sup>           | -                            |                             |                             |                             |                   |                   |      |       |
|                                | D       | .377 <sup>a</sup>           | .588 <sup>a</sup>            | -                           |                             |                             |                   |                   |      |       |
|                                | C       | .157 <sup>a</sup>           | .311 <sup>a</sup>            | .402 <sup>a</sup>           | -                           |                             |                   |                   |      |       |
|                                | M       | -.162                       | .095                         | -.110                       | .039                        | -                           |                   |                   |      |       |
| Wife                           | F       | <del>.758<sup>a</sup></del> | <del>.186<sup>a</sup></del>  | <del>.164<sup>a</sup></del> | <del>.228<sup>a</sup></del> | <del>-.125</del>            | -                 |                   |      |       |
|                                | T       | <del>.354<sup>c</sup></del> | <del>.814<sup>a</sup></del>  | <del>.562<sup>b</sup></del> | <del>.346<sup>b</sup></del> | <del>.084</del>             | .362              | -                 |      |       |
|                                | D       | <del>.085<sup>b</sup></del> | <del>-.083<sup>a</sup></del> | <del>.850<sup>a</sup></del> | <del>.252<sup>a</sup></del> | <del>.205</del>             | .014              | .220              | -    |       |
|                                | C       | <del>.217<sup>c</sup></del> | <del>.257<sup>a</sup></del>  | <del>.179<sup>a</sup></del> | <del>.512<sup>a</sup></del> | <del>.263</del>             | .583 <sup>b</sup> | .376 <sup>a</sup> | .132 | -     |
|                                | M       | <del>-.149</del>            | <del>.126</del>              | <del>.046</del>             | <del>-.035</del>            | <del>.991<sup>a</sup></del> | -.209             | .075              | .328 | -.339 |

<sup>a</sup>  $p < .01$ <sup>b</sup>  $p < .05$ <sup>c</sup>  $p < .10$ 

\*F = Savings and Investment

T = Travel Reservations

D = Dining Out

C = Shopping for an Automobile

M = Shopping at Supermarket



## REFERENCES

- Angelmar, Reinhard and Bernard Pras (1978), "Verbal Rating Scales for Multinational Research," European Research, 6 (March), 62-67.
- Bagozzi, Richard P. (1980), Causal Models in Marketing, New York: Wiley.
- Bishop, Yvonne M.M., Stephen E. Feinberg, and Paul W. Holland (1975), Discrete Multivariate Analysis, Cambridge, MA: M.I.T. Press.
- Brown, Rex V. (1969), Research and The Credibility of Estimates, Boston: Division of Research, Graduate School of Business Administration, Harvard University.
- Buzzell, Robert D. (1968), "Can You Standardize Multinational Marketing?" Harvard Business Review, 46 (November), 102-113.
- Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminate Validation by The Multitrait-Multimethod Matrix," Psychological Bulletin, 56 (March), 81-104.
- Cochran, W.G. (1968), "Errors of Measurement in Statistics," Technometrics, 10 (November), 637-666.
- Cook, Thomas D. and Donald T. Campbell (1979), Quasi-Experimentation, Chicago: Rand-McNally.
- Cohen, Jacob (1960), "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, 20 (Spring), 37-46.
- Cronbach, Lee J. (1951), "Coefficient Alpha and The Internal Structure of Tests," Psychometrika, 16 (September), 297-334.
- Davis, Harry L. (1971), "Measurement of Husband-Wife Influence in Consumer Purchase Decisions," Journal of Marketing Research, 8 (August), 305-312.
- Davis, Harry L. (1972), "Determinants of Marital Roles in A Consumer Purchase Decision," Working Paper 72-14, European Institute for Advanced Studies In Management, April 1972.
- Davis, Harry L. and Benny P. Rigaux (1974), "Perception of Marital Roles in Decision Processes," Journal of Consumer Research, 1 (June), 51-62.
- Davis, Harry L. (1976), "Decision Making Within the Household," Journal of Consumer Research, 2 (March), 241-260.
- Davis, Harry L. and E.K. Easton Ragsdale (1979), "The Newlywed Game Phenomenon: Application to New Produce Development," Paper presented at the TIMS/ORSA Meetings, New Orleans, April 30.
- Douglas, Susan P. and Yoram Wind (1978), "Examining Family Role and Authority Patterns: Two Methodological Patterns," Journal of Marriage and the Family, 40 (February).

Douglas, Susan P. (1979), "A Cross-National Exploration of Husband-Wife Involvement in Selected Household Activities," in William L. Wilkie, ed. Advances in Consumer Research, Vol. VI, Ann Arbor, MI: Association for Consumer Research, 364-371.

Farley, John U. and John A. Howard, eds. (1975), Control of "Error" in Market Research Data, Lexington, MA: D. C. Heath.

Ferber, Robert (1955), "On the Reliability of Responses Secured in Sample Surveys," Journal of the American Statistical Association, 50 (September), 788-810.

Feldt, Leonard S. (1965), "The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty," Psychometrika, 30, (September), 357-370.

Fleiss, Joseph L. (1975), "Measuring Agreement between Two Judges on The Presence or Absence of A Trait," Biometrics, 31 (September), 651-659.

Goodman, Leo A. and William H. Kruskal (1954), "Measures of Association for Cross Classifications," Journal of The American Statistical Association, 49 (December), 732-764.

Goldberger, Arthur S. (1971), "Econometrics and Psychometrics: A Survey of Commonalities," Psychometrika, 36 (June), 83-107.

Griliches, Zvi (1974), "Errors in Variables and Other Inobservables", Econometrika, 42 (November), 971-998.

Guilford, J.P. (1954), Psychometric Methods, New York: Wiley.

Haberman, Paul W. and Jack Elinson (1967), "Family Income Reported In Surveys: Husbands Versus Wives," Journal of Marketing Research, 4 (May), 191-194.

Hoffman, Lois (1960), "The Effects of Employment of Mothers on Parental Power Relations and the Division of Household Tasks," Marriage and Family Living, 22 (February), 27-35.

Joreskog, K.G. (1970), "A General Method For Analysis of Covariance Structures," Biometrika, 57, No. 2., 239-251.

Joreskog, K.G. (1971), "Statistical Analysis of Sets of Congeneric Tests," Psychometrika, 36 (June), 109-133.

Kraemer, Helena C. (1979), "Ramifications Of A Population Model For K As A Coefficient Of Reliability," Psychometrika, 44 (December), 461-472.

Lipstein, Benjamin (1975), "In Defense of Small Samples," Journal of Advertising Research, 15 (February), 33-42.

Lipstein, Benjamin (1975), "On The Limits Of Reliability In Social And Commercial Surveys," Proceedings, ESOMAR Conference on Quality In Research, European Society for Opinion and Marketing Research, Amsterdam, 17-39.

Lipstein, Benjamin (1977), "Are You Paying More and Getting Less?" Proceedings, 23rd Annual Conference, Advertising Research Foundation, New York, 59-62.

Lord, Frederic M. and Melvin R. Novick (1968), Statistical Theories of Mental Test Scores, Reading, MA: Addison-Wesley.

Malpass, Roy S. (1977), "Theory and Method in Cross-Cultural Psychology," American Psychologist, 32 (December) 1069-1079.

Mayer, Charles S. (1978), "Multinational Marketing Research: The Magnifying Glass of Methodological Problems," European Research, 6 (March), 77-84.

Pessemier, Edgar and Albert Bruno (1971), "An Empirical Investigation of the Reliability and Stability of Selected Activity and Attitude Measures," in David M. Gardner, ed. Proceedings, 2nd Annual Conference, Association for Consumer Research, 389-403.

Rao, Prakasa V.V., and Rao, V. Nandini (1979), "An Evaluation of the Bardis Familism Scale in India," Journal of Marriage and the Family, 41, (May), 417-421.

Schreiber, E.M. (1975-76), "Dirty Data in Britain and the U.S.A.: The Reliability of 'Invariant' Characteristics Reported in Surveys," Public Opinion Quarterly, 39 (Winter), 493-506.

Siegel, Sidney (1956), Nonparametric Statistics, New York: Mc-Graw-Hill.

Silk, Alvin J. (1977), "Test Retest Correlations And The Reliability Of Copy Testing," Journal of Marketing Research, 14 (November), 476-486.

Silk, Alvin J. and Manohar U. Kalwani (1979), "Measuring Influence In Organizational Purchase Decisions," Working Paper 1077-79, Sloan School of Management, Massachusetts Institute of Technology, August.

Warren, Richard D., Joan Keller White, and Wayne A. Fuller (1974), "An Error-In-Variables Analysis of Managerial Role Performance," Journal of the American Statistical Association, 69 (December), 886-893.

Wells, William D. (1975), "Psychographics: A Critical Review," Journal of Marketing Research, 12 (May), 196-213.

Whiting, W.M. John (1968), "Methods and Problems of Cross-Cultural Research," in G. Linzey and E. Aronson eds., The Handbook of Social Psychology, Volume 2, 2nd edition, Reading, Mass: Addison-Wesley, 693-728.





